

# Overview of Lattice Gauge Theory Infrastructure Project

With the encouragement of the Department of Energy, members of the United States lattice gauge theory community have been working together for nearly five years to create the computational infrastructure we need to study quantum chromodynamics (QCD). Our effort has reached a critical stage. We are ready to build our first multi-teraflop/s computer, and, with the support of a grant from the DOE SciDAC program, we have written software that will enable members of our community to use this and other terascale computers effectively. The advent of terascale computing, coupled with recent improvements in the formulation of QCD on the lattice, provides an opportunity to make major advances in QCD calculations that are of great importance to the experimental programs in high energy and nuclear physics. Lattice QCD is an international field, and this opportunity has been recognized by theoretical physicists in Europe, Japan and the United Kingdom. They have announced ambitious plans to construct multi-teraflop/s computers dedicated to the study of QCD during the current year. If the U.S. lattice gauge theory community is to participate in the next round of discovery in our field and build a world class research program in it, we must do so as well.

Our scientific objectives are to understand the physical phenomena encompassed by QCD, and to make precise calculations of the theory's predictions. This requires large scale numerical simulations within the framework of lattice gauge theory. Such simulations are necessary to solve fundamental problems in high energy and nuclear physics that are at the heart of the Department of Energy's large experimental efforts in these fields. Major goals of the experimental programs in high energy and nuclear physics on which lattice QCD simulations can have an important impact are to: 1) verify the Standard Model or discover its limits, 2) understand the internal structure of nucleons and other hadrons, and 3) determine the properties of hadronic matter under extreme conditions. Lattice QCD calculations are essential to research in all of these areas.

In the next several years, the largest impact of lattice gauge theory calculations on the experimental high energy physics program will occur in the effort to make precise tests of the Standard Model. A significant portion of the experimental program is now devoted to the determination of elements of the Cabibbo-Kobayashi-Maskawa (CKM) matrix. By obtaining constraints on these matrix elements through the study of a number of different processes, one hopes to probe the limits of validity of the Standard Model and discover new physics beyond it. However, experimental results do not in general determine the CKM matrix elements without lattice calculations of QCD effects. In a number of important instances, it is the uncertainties in the lattice calculations, rather than in the experimental results, that limit the precision of the Standard Model tests. In an accompanying paper entitled *A case study of the impact of increased computational resources on lattice calculations: Constraints on Standard model parameters*, we analyze in detail the computational resources needed to bring the lattice uncertainties down to or below the experimental ones. Here we briefly outline our conclusions.

The present status of the determination of elements of the CKM matrix is summarized in Fig. 1 [1]. The quantities referred to in this figure are the ratio of  $b \rightarrow u$  and  $b \rightarrow c$  coupling constants ( $|V_{ub}/V_{cb}|$ ), the  $K^0 - \bar{K}^0$  mixing strength,  $\epsilon_K$ , the  $B_d - \bar{B}_d$  mixing amplitude  $\Delta M_d$ , and the  $B_s - \bar{B}_s$  mixing amplitude  $\Delta M_s$ . The quantity "sin 2 $\beta$ " is determined from measurement of CP-violation in the decay  $B \rightarrow J/\psi K_s$ . The parameters  $\bar{\rho}$  and  $\bar{\eta}$  are the least well known elements of the CKM

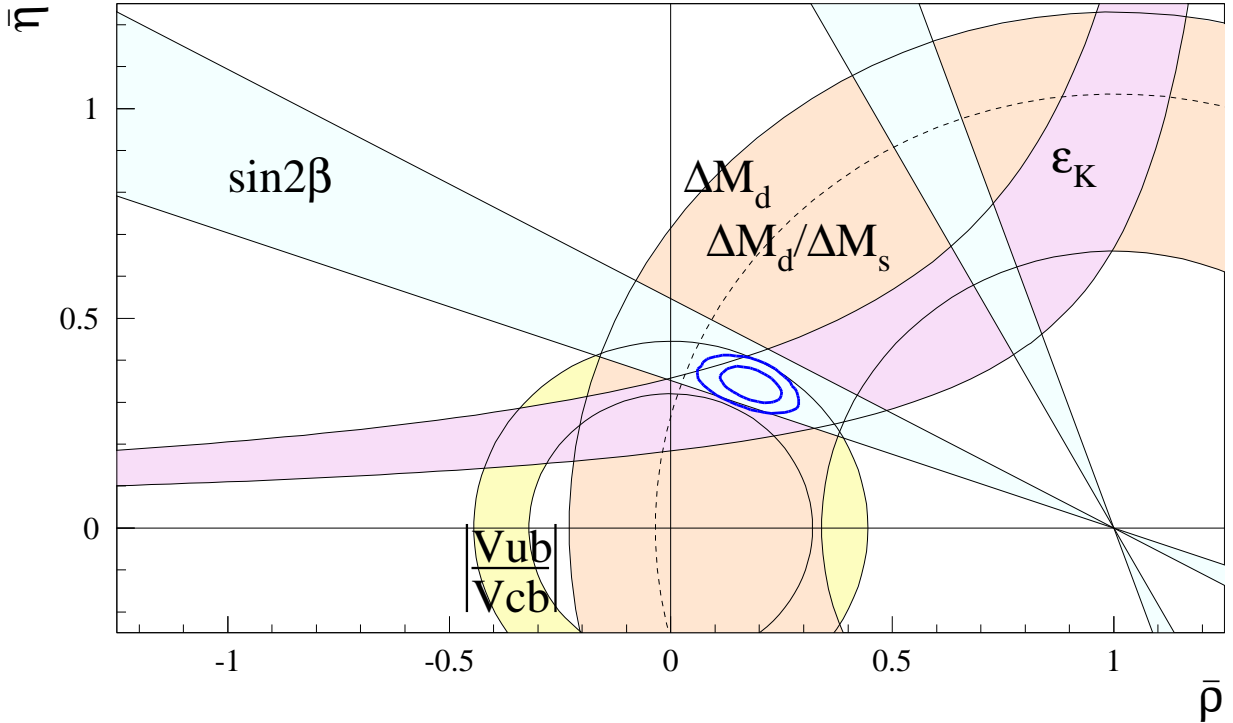


Figure 1: Allowed regions for  $\bar{\rho}$  and  $\bar{\eta}$ . The contours at 68% and 95% probability are shown. The full lines correspond to the 95% probability constraints given by the measurements of  $|V_{ub}|/|V_{cb}|$ ,  $\epsilon_K$ ,  $\Delta M_d$  and  $\sin 2\beta$ . The dotted curve bounds the region selected at 95% by the lower limit on  $\Delta M_s$ .

matrix. Each of the measured quantities constrains  $\bar{\rho}$  and  $\bar{\eta}$  as shown, with the error bands resulting from a combination of experimental and theoretical uncertainties. At present, all determinations are consistent with the allowed region shown by the ovals in the figure.

It is clear from this figure that a substantial reduction in the error band from  $\Delta M_d$  has the greatest potential to test the Standard Model. Next in potential is the  $\epsilon_K$  band. In both these quantities the lattice error is dominant. Once  $\Delta M_s$  is measured, a similar comment may apply there.

The impact of terascale computers on the determination of CKM matrix elements is illustrated in Table 1. The non-lattice uncertainties listed in the fourth column of this table are primarily experimental in origin. The last three columns show our estimates of the lattice errors that would be obtained with a computer that sustained 0.6, 6.0 and 60.0 teraflop/s for one year (TF-Yr), respectively. Uncertainties in the lattice calculations do not decrease as the inverse square root of the computing resources, because they are not dominated by statistics. The major sources of uncertainty are the extrapolation of the up and down quark masses to the chiral limit, and the extrapolation to the continuum limit. The relative importance of these factors depends on the quantity in question and the computing resources available, as is discussed in the accompanying document. Note that the error estimates in the 6.0 and 60. TF-Yr columns assume that two-loop perturbation theory calculations for matching lattice to continuum results will have been completed. The impact of decreasing lattice errors on the precision of Standard Model tests is illustrated in Fig. 2.

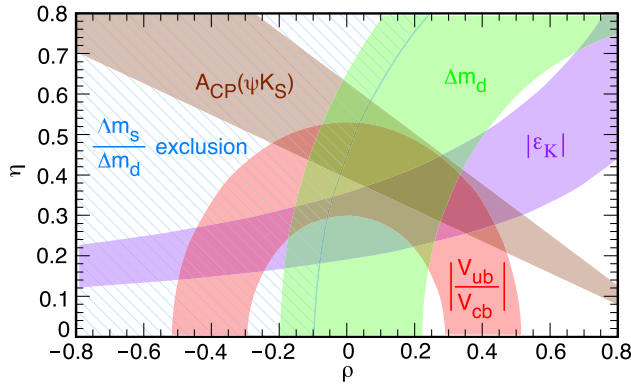
Measurement	CKM Matrix Element	Hadronic Matrix Element	Non-Lattice Errors	Current Lattice Errors	Lattice Errors 0.6 TF-Yr	Lattice Errors 6.0 TF-Yr	Lattice Errors 60. TF-Yr
$\epsilon_K$ ( $\bar{K}K$ mixing)	$\text{Im} V_{td}^2$	$\hat{B}_K$	10%	20%	12%	5%	3%
$\Delta M_d$ ( $\bar{B}B$ mixing)	$ V_{td} ^2$	$f_{B_d}^2 B_{B_d}$	6%	30%	16%–26%	8%–10%	6%–8%
$\Delta M_d/\Delta M_s$	$ V_{td}/V_{ts} ^2$	$\xi^2$	—	12%	8%	6%	3%–4%
$B \rightarrow (\rho) l\nu$	$ V_{ub} ^2$	$\langle \rho   (V - A)_\mu   B \rangle$	7%	15%	10%–13%	5.5%–6.5%	4%–5%
$B \rightarrow (D^*) l\nu$	$ V_{cb} ^2$	$ \mathcal{F}_{B \rightarrow (D^*) l\nu} ^2$	2%	4.4%	3%–4%	1.8%–2%	1%–1.4%

Table 1: Impact of lattice QCD on the determination of CKM matrix elements. The non-lattice errors in the fourth column are primarily experimental. The last three columns show the improvements in lattice errors that we estimate would be obtained with computers sustaining 0.6 and 6.0 and 60. Tflops for one year.

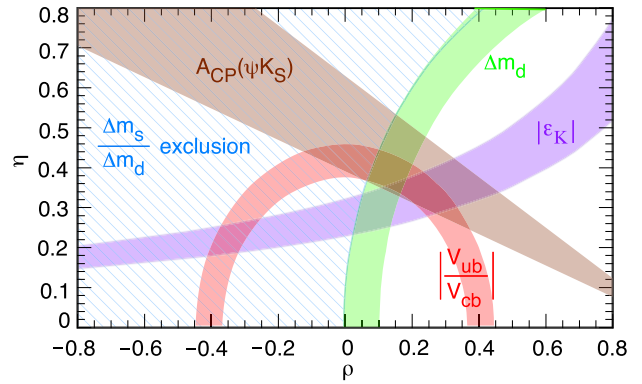
The 0.6 TF-Yr calculation is in progress. It is the starting point for our error analysis, and it has already yielded determinations of the strong coupling constant,  $f_\pi$ ,  $f_K$ , and mass splittings in the onium spectra that agree with experiment within statistical and systematic errors estimated to be 3%, as is illustrated in Fig. 3 [2].

It should be emphasized that the calculations envisioned in the last three columns of Table 1 and those shown in the right hand column of Fig. 3 fully take into account the effects of up, down and strange sea quarks, the only three quarks light enough so that their vacuum polarization has a substantial impact on the results. Thus, these are full QCD calculations. In full QCD calculations, the overwhelming fraction of the computer resources are taken up generating representative lattice gauge configurations, which are snapshots of the QCD vacuum. The same lattices can be used to evaluate all of the matrix elements listed in Table 1 as well as many other physical quantities of interest. The calculations of errors in the last three columns of Table 1 are based on the use of improved staggered quarks. Our collaboration is actively investigating other approaches, including domain wall/overlap quarks which have nearly exact chiral symmetry on the lattice, and may yield even smaller errors. The inclusion of these quarks will allow important checks of the systematic errors, as well as allow the reliable calculation of additional quantities. Although we have focused on the impact of lattice calculations on the determination of CKM matrix elements, we should emphasize that they will also be important for studies of the internal structure of hadrons, and of hadronic matter under extreme conditions. Finally, if, as expected, the Standard Model turns out to be the low energy limit of a more fundamental theory, then that theory will very likely have a strong coupling regime that generalizes QCD. We therefore believe that large scale numerical simulations will continue to play an important role in the study of high energy physics for the foreseeable future.

CKM today ...



... and with 2–3% theory errors.



And with B Factories ...

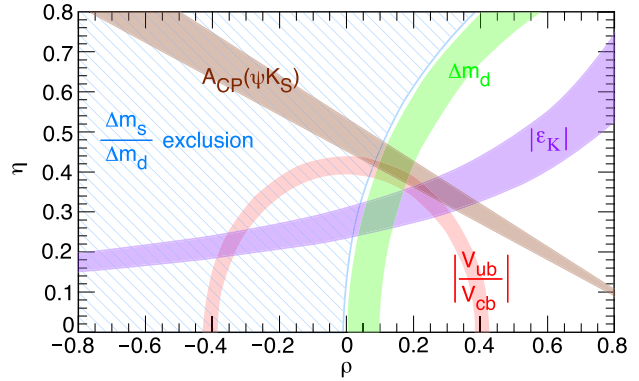
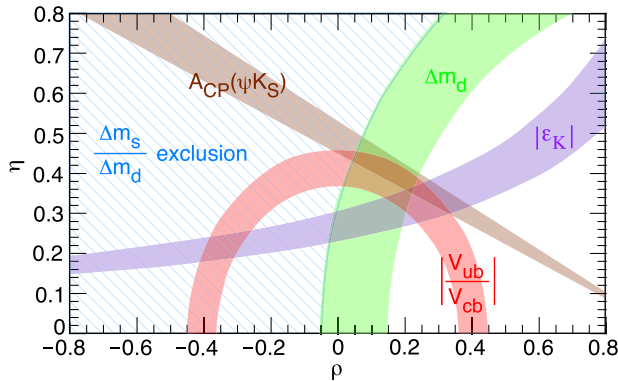


Figure 2: The impact of the B factories and improvements in lattice calculations on parameters of the CKM matrix as estimated by the CLEO-c collaboration (2001).

The 0.6 TF-Yr calculation is straining the resources available to lattice gauge theorists at the DOE and NSF supercomputer centers despite the fact that our community is one of, if not the largest users of these facilities. If the computational resources available at these centers grow with Moore's law, it would be over five years before the 6.0 TF-Yr calculation could be completed at them. However, by taking advantage of simplifying features of lattice QCD calculations, such as regular grids and uniform, predictable communications between processors, it is possible to construct computers for lattice QCD that are far more cost effective than general purpose supercomputers, which must perform well for a wide variety of problems including those requiring irregular or adaptive grids, non-uniform communication patterns, and massive input/output capabilities. In addition, lattice QCD calculations require significantly less memory than most large scale applications, which also serves to reduce the cost of computers dedicated to our field relative to those that must serve a broad range of disciplines. The construction of special purpose computers for lattice gauge theory was pioneered by the lattice group at Columbia University, and has been vigorously pursued by lattice gauge theorists in other countries.

We are pursuing two hardware tracks. One is the QCD on a Chip (QCDOC), which was designed

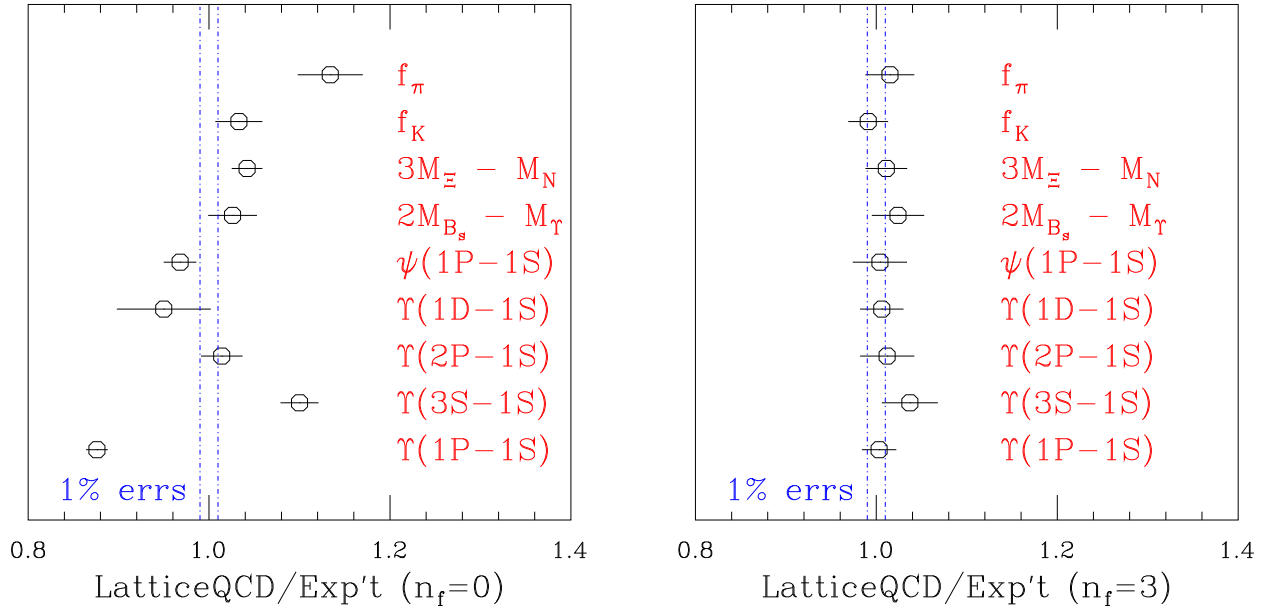


Figure 3: Lattice QCD results divided by experimental results for nine different quantities, without (left panel) and with (right panel) quark vacuum polarization. These quantities were chosen because one expected *a priori* to have small systematic errors. Obtaining similar accuracy for the quantities in Table 1 is significantly more challenging.

by lattice gauge theorists at Columbia University in collaboration with colleagues at IBM. The other is commodity clusters optimized for the study of QCD, which are being built and tested at Fermi National Accelerator Laboratory (FNAL) and Thomas Jefferson National Accelerator Facility (JLab) under our SciDAC grant. The initial components of the QCDOC have been thoroughly tested, and the designers are ready to begin construction of terascale machines. Tests with a gate level simulator and with the initial hardware indicate that the QCDOC will meet its design goal of \$1 per *sustained* megaflop/s. By comparison, the most powerful machines at NERSC and in the NSF PACI Program, as well as the heralded Japanese Earth Simulator have price performances in the range of \$20 to \$50 per sustained megaflop/s for lattice QCD simulations. As a result, in 2004 the QCDOC architecture offers substantial cost performance advantages for QCD. Clusters built from carefully selected commodity components are expected to provide even better price performance for single systems as large as a few sustained teraflop/s by 2005. Therefore, we propose to build a 10 teraflop/s QCDOC at Brookhaven National Laboratory in 2004, and to construction large scale clusters at FNAL and JLab in 2005 and 2006. It is only by this sustained investment in large-scale, cost effective resources that the lattice QCD community can fulfill its potential to contribute to to research in high energy and nuclear physics.

Lattice gauge theorists in Europe, Japan and the United Kingdom are moving aggressively to build computational resources on the scale we have proposed. The APE group, which is centered in Italy, has a long history of building special purpose computers for the study of QCD. It is about to begin production of its latest machine, the apeNEXT, which has performance and price performance characteristics similar to the QCDOC. The German lattice gauge theory community will obtain apeNEXT computers sustaining a total of twelve teraflop/s in 2004, while in the same time

period the Italian community will obtain apeNEXTs capable of sustaining several teraflop/s. The British lattice gauge theory consortium, UKQCD, and the Japanese funded Riken/BNL Research Center have each funded five teraflop/s sustained QCDOCs which will be delivered in mid-2004.<sup>1</sup> Japanese lattice gauge theorists have begun use of the Earth Simulator, and have very ambitious plans for upgrading their current facilities, which already sustain approximately one teraflop/s. Thus, despite the investments that the DOE has made in the development of the QCDOC and in cluster prototyping, the U.S. lattice gauge theory community will be left far behind its international competitors if we do not begin construction of the QCDOC immediately. What is at stake is not simply the ability to carry out the near term research program outlined above, but the viability of the field in the United States. It is very difficult to see how outstanding young people can be attracted to the field, or those in it continue, if the computational resources available to the U.S. community lag those in other countries by an order of magnitude or more. Lattice gauge theory was invented in the United States, and U.S. physicists have traditionally been among the intellectual leaders of the field. It would be a major blow if that leadership were lost just as the field is entering an exciting period of discovery.

Our project was reviewed in February 2003 by a panel of physicists and computer scientists chaired by Frank Wilczek. Their report is reproduced in full in an accompanying document. Among the major points made in the report were:

- “The scientific merit of the suggested program is very clearly outstanding.”
- “The proposed programs are of considerable interest from the point of view of computational science, since they could provide convincing models and demonstrations of the use of cost effective special architectures for scientific problems.”
- “It is prudent, as well as interesting, to pursue both [hardware] tracks”
- “The software development component of the proposal is also novel . . . and extremely important.”
- “Sustained operation at the multi-teraflop level is both necessary and sufficient for the U.S. effort to match existing European and Japanese initiatives in the immediate future.”
- “Both the proposers and the DOE should recognize that this is an endeavor that is not likely to be exhausted in 4 years or even in 10.”

We call particular attention to the last bullet. Our community needs sustained funding for computational infrastructure if we are to meet our scientific objectives and build an internationally competitive research program. We expect to receive continued funding from the DOE SciDAC Program, and hope to receive new funding from the Offices of Nuclear Physics and Advanced Scientific Computing Research. For example, the Nuclear Science Advisory Committee’s Nuclear Theory Subcommittee has recommended that “even if external sources cannot be secured and the funding is flat, a minimal sum of \$3M/year from the nuclear science budget should be allocated

---

<sup>1</sup>While the Riken QCDOC will be available to physicists at Columbia and BNL, it will not be open to other members of the U.S. lattice gauge theory community, and therefore cannot serve as the basis for a national program.

to hardware investment.” [3] However, if we are to be successful, it is essential that the Office of High Energy Physics play a decisive leadership role, both in funding the QCDOC now when it will have maximal impact, and in providing sustained support for lattice QCD.

We have set out a plan that would place the U.S. lattice gauge theory community in a leadership position in this important area. To carry out this plan requires resources beyond those that could reasonably be supplied by the Office of High Energy Physics alone, and we are vigorously seeking additional support from other sources. However, since funding from other sources is not assured, we have been asked what the minimum funding from High Energy Physics would need to be to ensure a strong program. From Table 1 and the announced plans of other countries we conclude that at a minimum a hardware budget of \$3M per year beginning in FY 2004 is required. In order to carry out the 6.0 TF-YR calculations in a timely fashion and to remain internationally competitive, we must at the very least build a 5 teraflop/s QCDOC this year, and begin building a substantial cluster in 2005. During this period the QCDOC will be essential for the generation of lattices and the extraction of physics from them, while the clusters will be critical for perturbative calculations and important parts of the physics analysis. Although the cost for construction of the QCDOC could be spread over the 2004 and 2005 fiscal years, continued funding at the \$3M level is essential to permit us to exploit Moore’s law, and work towards capability in the 60–100 teraflop/s regime before the end of the decade.

## References

- [1] M. Ciuchini, E. Franco, F. Parodi, V. Lubicz, L. Silvestrini and A. Stocchi, arXiv:hep-ph/0307195.
- [2] C.T.H. Davies, *et. al.*, arXiv:hep-lat/0304004, Phys. Rev. Lett. **92**, 022001 (2004).
- [3] J. Carlson, *et. al.*, arXiv:nucl-th/0311056, Page 43.